

6.1 负载均衡

2022年3月15日 19:42

通常情况下，**负载均衡**可以分为两种：

- 一种是请求负载均衡，即将用户的请求均衡地分发到不同的服务器进行处理；
- 另一种是数据负载均衡，即将用户更新的数据分发到不同的存储服务器。

	优点	缺点	适用场景	典型应用/系统
轮询策略	实现简单，服务器负载均衡，可解决服务器节点异构的问题	1. 处理每次请求的服务器不确定，不适合有状态请求的场景 2. 没考虑请求开销不同造成的不均衡问题	适用于用户请求所需资源比较接近的场景，以及无状态请求场景	Nginx
随机策略	实现简单，服务器负载基本均衡	1. 处理每次请求的服务器不确定，不适合有状态请求的场景 2. 没考虑服务器节点异构性和请求开销不同造成的不均衡问题	适用于集群中服务器节点处理能力相差不大，用户请求所需资源比较接近的无状态请求的场景	Dubbo
哈希和一致性哈希策略	1. 哈希函数设置合理，服务器负载均衡，且相同key的请求可落在同一个服务器节点，适合有状态请求的场景 2. 带虚拟节点的一致性哈希策略，可以解决服务器节点异构的问题	1. 实现相对复杂 2. 没考虑请求开销不同造成的不均衡问题	适用于用户请求所需资源比较接近的场景，也适用于有状态请求的场景	Redis、Memcache、Cassandra